



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Cumulative cultural evolution in the laboratory

Citation for published version:

Kirby, S, Cornish, H & Smith, K 2008, 'Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language', *Proceedings of the National Academy of Sciences (PNAS)*, vol. 105, no. 31, pp. 10681-10686. <https://doi.org/10.1073/pnas.0707835105>

Digital Object Identifier (DOI):

[10.1073/pnas.0707835105](https://doi.org/10.1073/pnas.0707835105)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Proceedings of the National Academy of Sciences (PNAS)

Publisher Rights Statement:

This is an author's accepted manuscript of the following article: Kirby, S., Cornish, H. & Smith, K. Aug 2008, "Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language", in *Proceedings of the National Academy of Sciences of the United States of America - PNAS*. 105, 31, p. 10681-10686, © 2008 by The National Academy of Sciences of the USA. The final publication is available at www.pnas.org/content/105/31/10681

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Cumulative Cultural Evolution in the Laboratory: an experimental approach to the origins of structure in human language

Simon Kirby^{1*}, Hannah Cornish¹ & Kenny Smith²

¹School of Philosophy, Psychology & Language Sciences, University of Edinburgh, 40 George Square, Edinburgh, UK, EH8 9LL

²Division of Psychology, Northumbria University, Northumberland Building
Northumberland Road, Newcastle-upon-Tyne, UK, NE1 8ST

*To whom correspondence should be addressed. Email: simon@ling.ed.ac.uk. Fax: +44 131 650 3962

Classification: Social Sciences (Psychology) and Biological Sciences (Evolution)

Abstract

We introduce an experimental paradigm for studying the cumulative cultural evolution of language. In doing so we provide the first experimental validation for the idea that cultural transmission can lead to the appearance of design without a designer. Our experiments involve the iterated learning of artificial languages by human participants. We show that languages transmitted culturally evolve in such a way as to maximize their own transmissibility: over time, the languages in our experiments become easier to learn and increasingly structured. Furthermore, this structure emerges purely as a consequence of the transmission of language over generations, without any intentional design on the part of individual language learners. Previous computational and mathematical models suggest that iterated learning provides an explanation for the structure of human language, and link particular aspects of linguistic structure with particular constraints acting on language during its transmission. The experimental work presented here shows that the predictions of these models, and models of cultural evolution more generally, can be tested in the laboratory.

Cumulative Cultural Evolution in the Laboratory: an experimental approach to the origins of structure in human language

Introduction

The emergence of human language has been cited by Maynard Smith and Szathmary (1) as the most recent of a small number of highly significant evolutionary transitions in the history of life on earth. The reason they give for language's inclusion in this list is that it enables an entirely new system for information transmission: human culture. Language is unique in being a system that supports unlimited heredity of cultural information, allowing our species to develop a unique kind of open-ended adaptability.

While this feature of language as a carrier of cultural information is obviously important, we have argued that there is a second sense in which language is an evolutionary milestone: each utterance has a dual purpose, not only carrying semantic content, but also conveying information about its own construction (2-5). Upon hearing a sentence, a language learner will use the structure of that sentence to make new inferences about the language that produced it – this allows learners to reverse engineer the language of their speech community from the utterances they hear. Language is thus both a conveyer of cultural information (in Maynard Smith and Szathmary's sense) and itself culturally transmitted. This makes language an evolutionary system in its own right (2-3), suggesting a new approach to the explanation of linguistic structure. Crucially, language also represents an excellent test domain for theories of cultural evolution in general since it is relatively well understood, both in terms of its acquisition and processing, and has interesting non-trivial, yet well documented, structure.¹

Over the past ten years there have been a wide range of computational and mathematical models which have looked at a particular kind of cultural evolution termed **iterated learning** (4-13):

Iterated Learning: a process whereby some behaviour is acquired by an individual by observing a similar behaviour in another individual who acquired it in the same way.

Spoken (or signed) language is an outcome of iterated learning. Although there may be some circumstances where aspects of language are explicitly taught, acquired from a written form, or arise from deliberate invention, almost all of the features of the languages we speak are the result of iterated learning. Models of this process (4-13) demonstrate that, over repeated episodes of transmission, behaviours transmitted by iterated learning tend to: 1. become easier to learn, and 2. become increasingly structured. Note that this is a cumulative process, and is not considered to arise from the explicit intentions of the individuals involved. Rather, this type of cultural evolution is an “invisible hand” process leading to phenomena that are the result of human action but are not intentional artefacts (14).

While these models are indicative of the power of cultural evolution in explaining language structure, there remains scepticism as to how well computational models of learning match the abilities and biases of real human learners. For example, responding to a growing body of computational models of the emergence of multi-word utterances from unstructured randomness (5,8,10,11,15), Bickerton notes, “Powerful and potentially interesting though this approach is, its failure to incorporate more realistic conditions (perhaps because these would be more difficult to simulate) sharply reduces any contribution it might make towards unraveling language evolution. So far, it is a classic case of looking for your car-keys where the street-lamps are.” (16, p522).

¹ From a practical perspective it is also an ideal subject for study in being relatively straightforward to precisely record and analyse.

What is needed, therefore, is an experimental paradigm for studying the evolution of complex cultural adaptations using real human participants. Ideally, this paradigm should mirror previous computational and mathematical models and provide a test for the claim that iterated learning leads to adaptively structured languages. It should demonstrate whether cumulative adaptive evolution without intention is possible purely by virtue of cultural transmission.

In this paper, we implement such a paradigm and demonstrate for the first time cumulative, adaptive, non-intentional cultural evolution of an artificial language in a laboratory population of human participants.

Diffusion chains

Diffusion chain studies provide the best example of experimental treatments of iterated learning. In these experiments a participant observes some target behaviour (provided by the experimenter), and is then required to replicate that behaviour in some way which can be observed by a second participant. This second participant in turn attempts to replicate the first participant's behaviour for a third participant, and so on (we'll refer to each iteration of this cycle as one **generation**). Using this procedure, we can observe the diffusion of behaviour through a chain of cultural transmission. The first reported use of this methodology was Bartlett's in 1932 (17), but it was not until recently that researchers began to apply this approach systematically (18-24).

The most recent, and arguably most significant, instance of a diffusion chain experiment is the work of Horner *et al.*, which explores the cultural transmission of tool-use strategies in populations of chimpanzees and children (24). Diffusion chains are set up in which an experimenter demonstrates one of two possible techniques for opening a puzzle box ("artificial fruit") to a participant. Subsequent participants observe their predecessor's box-opening behaviour and then in turn become the model for the next generation. These experiments demonstrate clearly that both chimpanzees and children are capable of high-fidelity cultural transmission – the box-opening technique used by the last participant in the chains (of up to 10 individuals) is the same as that demonstrated to the first participant, with a chain of faithful transmission between the two.

While these experiments are ground-breaking, particularly in showing that cultural transmission can be studied empirically even in species other than humans, they do not support our claim that culture leads to cumulative non-intentional adaptation. This is because the behavioural information which is being transmitted is drawn from a limited set of possibilities. For example, in the puzzle-box study, there are essentially two different strategies for opening the box. There is simply not enough complexity in the task to demonstrate adaptation, let alone cumulative adaptation. In any case, both strategies seem to be equivalently "adaptive" in cultural and environmental terms, in that they both open the box, and are both transmittable.

To get around these problems and to allow us to make a direct comparison with human language, we replicate the basic diffusion-chain design, but with a more complex **artificial language learning** task (25-26): that of labelling visual stimuli with strings of written syllables. In order to make this task tractable, we use adult human participants and observe the cultural evolution of the artificial language for ten cultural generations.

This work bears some resemblance to a recent body of experimental work on the shared construction of communication systems (27-30). Of particular relevance is a recent paper by Selten and Warglien (30), which demonstrates that pairs of participants can (sometimes) create structured and efficient communication systems over the course of repeated interaction. The major difference between the experiments described here and this work is the role of intentional design. In Selten and Warglien's experiments, as with those of Galantucci (27) and Garrod *et al.* (28-29), participants interact repeatedly with the explicit goal of arriving at a shared system for communication – the sys-

tems they construct are therefore the outcome of conscious design. Our diffusion chain experiment allows us to explore whether structured languages can emerge without intentional design, as has been argued to be the case for language (14).

Experiment 1 design

Participants are asked to learn an “alien” language made up of written labels for visual stimuli. The stimuli are pictures of coloured objects in motion, and the labels are sequences of lower-case letters (see Fig. 1 for an example, and the Methods section for more details).

For training purposes, the language to be learned (a set of string-picture pairs) is divided randomly into two sets of approximately equal size: the SEEN set and the UNSEEN set. A participant is trained on the SEEN set, being repeatedly presented with each string-picture pair in random order (see Methods for more details on the training regime). During subsequent testing, the participant is presented with a picture and asked to produce the string they think the alien would give for that picture. Participants are tested on both the SEEN and UNSEEN sets in their entirety.

The initial set of labels in the language are generated and assigned randomly, and the first participant in the experiment is trained on this random language. Subsequent participants are trained on the output of the (final) testing of the previous participant, which is re-divided into new SEEN and UNSEEN sets. Note that the experimental procedure is equivalent for all participants despite the different sources of training data: at no stage are participants told that they are being trained on the output of another person, nor did any guess this was the purpose of the experiment. Crucially, participants believe they are copying the input language as best they can – a post-test questionnaire revealed that many participants did not even realise that they were being tested on stimuli they had not seen in training. This makes intentional design on the part of the participants unlikely – to put it another way, the participants’ goal is to reproduce the language, not improve it in some way (we return to this point in the discussion section).

Our hypothesis is that we will observe cumulative adaptive evolution of the language being transmitted in this experiment – we should see the emergence of adaptive structure in response to the pressure on the language to be faithfully transmitted from generation to generation. If this hypothesis is correct we should see two things: 1. an increase in the learnability of the language over generations (i.e., a decrease in transmission error), and 2. the evolution of linguistic structure (i.e., an increase in predictability in the mapping between meanings and signals).

In order to test this, we devised two measures. Firstly, we used a measure of string similarity to compare words in the languages of participants at adjacent generations (see Methods). The Levenshtein edit distance between pairs of words (i.e., the smallest number of character insertions, replacements and deletions required to transform one word to the other) provides a reasonable theory-neutral measure of distance. We normalised this for length of words such that identical strings have a distance of 0, and maximally distinct ones have a distance of 1. The mean distance between all the words in a participant’s output and the corresponding words in the previous generation’s output gives a straightforward measure of the error in transmission of the language.

Secondly, we constructed a novel measure of linguistic structure, based on measures of compositionality used in some computational models (12). Our aim is to quantify the degree to which the mapping between meanings (visual scenes) and signals (character strings) is **systematic** – an obvious hallmark of structure in human language. A language is systematic if patterns of similarity and dissimilarity in signals provide information about the relationship between the meanings those signals map on to. Accordingly, we calculated the correlation between all pairs of edit-distances in the set of signals and the corresponding distances between meanings (i.e., whether they differed on shape, colour and/or movement). Using Monte-Carlo techniques, it is possible to calculate the ex-

tent to which this alignment between meaning and signal differs from the alignment we would expect to see by a random, unstructured assignment of signals to meanings (see Methods for details).

Experiment 1 results

The results of our first experiment, involving four separate diffusion chains of ten participants each, are shown in Fig. 2. Each of these chains was initialized with a different random language. There is a clear and statistically-significant decrease in transmission error between the initial and final generations (mean decrease of 0.748, standard deviation of 0.147, $t(3)=8.656$, $p < 0.002$). This confirms the first of our two predictions above: the language is adapting to become increasingly transmissible from generation to generation. Indeed, towards the end of some chains the language is transmitted perfectly – these participants produced exactly the same strings for every meaning as their predecessor, despite the fact that they received no exposure to the strings associated with half of those meanings.

How is this possible? Is there any structural evolution of the language taking place (the second of our two predictions)? As Table 1 shows, the number of distinct strings in each language decreases rapidly. The initial random languages are completely unambiguous: every meaning is expressed by a distinct signal. The transmission process cumulatively introduces ambiguity as single strings are re-used to express more and more meanings. In other words, the languages gradually introduce **underspecification** of meanings. Clearly, the reduction in the number of strings must make a language easier for participants to learn, but it cannot on its own account for the results we see. For example, it does not explain how, in some chains, participants are able to produce the correct signal for *every* meaning, including meanings drawn from the UNSEEN set.

The answer to this puzzle lies in the structure of the languages. The initial random language is, by definition, unstructured: nothing in the set of signals gives any systematic clue to the meanings being conveyed. To put it another way, the only way to learn this language is by rote. Equally, if a language is randomly underspecified, then rote learning is the only way it can be acquired. For example, if the same signal is used for a black spiralling triangle and a red bouncing square then a learner must see this signal used for both of these meanings in order to learn it. Because we deliberately hold items back from the SEEN set, rote learning for all meanings is impossible. For learners to be able to successfully generalize to unseen meanings, there must be **systematic underspecification**.

We can observe exactly this kind of structure evolving by examining a language as it develops in the experiment. For example, by generation 4 in one of the diffusion chains the string *tuge* is used exclusively for all pictures with an object moving horizontally. The distribution of the other strings in the language is more idiosyncratic and unpredictable at this stage. By generation 6, *poi* is used to refer to most spiralling pictures, but there are exceptions for triangles and squares. Blue spiralling triangles or squares are referred to as *tupin*, and red spiralling triangles or squares are *tupim*. In the following generation, these exceptional cases are reduced to the blue spiralling triangle and the red spiralling square. By generation 8 (shown in Fig. 3), and also for generations 9 and 10, the language has settled on a simple system of regularities whereby everything that moves horizontally is *tuge*, all spiralling objects are *poi*, and bouncing objects are divided according to shape (see Fig. 3).

It is precisely because the language can be described using this simple set of generalisations that participants are able to correctly label pictures that they have never previously seen. This directly ensures the stable cultural transmission of the language from generation to generation, despite the incomplete training data that each learner of the language is exposed to.

Our structure measure confirms that the languages evolve to become more structured. As can be seen in Fig. 2b, significantly non-random structure in the mapping from meanings to signals rapidly emerges. Furthermore, the languages produced by the final generation are significantly more struc-

tured than the initial languages (mean increase of 5.578, standard deviation of 2.968, $t(3)=3.7575$, $p<0.02$).

Languages in this experiment are evolving to be **learnable**, and they are doing this by becoming **structured**. This confirms our hypothesis regarding the cultural evolution of language. However, we are interested in whether it would be possible for a language to evolve that was learnable and structured, but also **expressive**. In other words, a language that would be able to label meanings unambiguously. Such a language would not be able to rely on systematic underspecification of meanings, but must rather find some other means of gaining structure.

Experiment 2 design

Accordingly, in the second experiment we made a single minor modification: we “filtered” the SEEN set before each participant’s training. If any strings were assigned to more than one meaning, all but one of those meanings (chosen at random) was removed from the training data. This effectively removes the possibility of the language adapting to be learnable by introducing underspecification – filtering ensures that underspecification is an evolutionary dead-end. This filtering process, although artificial, is an analog of a pressure to be expressive that would come from communicative need in the case of real language transmission.

Experiment 2 results

As expected, under the modified regime, the overall number of words in participants’ output remains comparatively high throughout the experiment, as shown in Table 2. Fig. 4a shows how transmission error changes as the language evolves. Once again, it is clear that the languages are becoming more learnable over time (mean decrease of 0.427, standard deviation of 0.106, $t(3)=8.0557$, $p<0.002$) despite being unable to introduce the kind of underspecification we saw before. Furthermore, it is clear from Fig. 4b that the languages are becoming increasingly structured over time, just as before (mean increase of 6.805, standard deviation of 5.390, $t(3) = 2.525$, $p<0.05$). Since filtering rules out the generalisations that emerged in the previous experiment, a different kind of structure that does not rely on underspecification must be emerging.

If we examine the languages at particular stages in their cultural evolution, we can see exactly what this structure is. For example, Fig. 5 shows the language output by a participant at generation 9 in one of the diffusion chains. Looking at this language it becomes immediately clear that there is structure within the signals. We can analyse each signal as three **morphemes** expressing colour, shape, and movement respectively, with one exceptional irregularity (*renana* for bouncing red circle). It turns out that this general structure emerges by at least generation 6 and persists to the end of the experiment, although the details change as some morphemes are lost, or reanalyzed from generation to generation (see Supporting Information for the complete set of languages).

Discussion

What we have observed here, for the first time in laboratory conditions, is cumulative cultural adaptation² without intentional design. Just as previous computational models have predicted (4-13), the culturally-evolving language has adapted in such a way as to ensure its successful transmission from generation to generation, despite the existence of a bottleneck on transmission imposed by the incomplete exposure of each participant to the language. Cultural adaptation results in languages

² Recall that we expect adaptation with respect to learnability and structure, but not necessarily expressivity. Cumulative adaptation therefore does not imply that the languages will become more functional with respect to communication.

that circumvent this transmission problem by exploiting structure in the set of meanings to be conveyed.

We have shown in all of our experiments that languages, by virtue of being culturally transmitted, become increasingly learnable, and increasingly structured. An obvious question is: to what extent does the structure we see emerging resemble structures found in real human languages?

In the first experiment, we saw underspecification introduced into the language. This underspecification was not random, but systematic in that similar meanings were given the same label. The form of the language reflected regularities in the visual scenes, namely that they consisted of shape, colour and motion. Of course, in the experiment this process ran unchecked, and in some cases led to languages where almost every meaning was expressed by a single signal.

The languages in our first experiment could therefore be seen as counter-functionally ambiguous. However, there is another way of thinking about our results. Rather than seeing the emerging language as ambiguous, some participants thought it revealed something about the way the aliens saw the world. For example, in post-test discussions, one participant noted that “colour is not important to these aliens”. This suggests that the participants did not consider the language to be ambiguous as such, but rather that it reflected the meaning distinctions that the aliens were interested in communicating. The collapse of distinctions based on colour (which eventually occurred in all four replications of the first experiment), in favour of distinctions based on shape and movement, is compatible with the literature on a shape bias, an expectation that words will refer to shapes of objects, rather than properties of objects such as colour or texture (31). It may be that while adapting to be learnable by eliminating semantic distinctions, the languages in the experiment retains those distinctions which seem most salient and/or likely to be labeled linguistically.

Systematic underspecification similar to that found in the experiments is an important feature of natural language. For example, in the class of nouns only proper names refer to specific entities. Others are underspecified and typically correspond to natural classes. However, this is not the only way in which the structure of the set of meanings we convey makes itself felt in linguistic expressions. Most obviously, natural languages exhibit the species-unique property of **compositionality** in syntax and morphology.³ The meaning of an expression is normally a function of the meanings of sub-parts of that expression, and the way they are put together. It is precisely this property that we hypothesise allows language to be both learnable and expressive.

Expressivity in human language is presumably a consequence of the fact that language is used for communication, and may also be attributable to predispositions of child language learners (32-33). In one computational model of iterated learning (8), an expressivity requirement is simply enforced by filtering out ambiguous meaning-string examples from the data given to the learner, leaving a training set with a unique one-to-one mapping between meanings and strings: although learners would still be free to infer ambiguous strings, such ambiguity would not be transmitted to the following generation.

We implemented exactly this filtering process in the second experiment, to dramatic effect. This is despite the fact that, to the participants, the conditions in this experiment were essentially identical to those in the previous one – as before, after being presented with string-picture pairs they had to recall these and generalise to unseen pictures. Nevertheless, unlike in the previous experiment, systematic compositional structure emerged. Rules evolved for constructing signals out of a combination of meaningful sub-strings, and these rules tended to be transmitted from generation to genera-

³ Arguably, the dance of honey bees (34) and the calls of Campbell’s monkeys (35) are both minimally compositional. However, there is no evidence (as yet) for culturally-transmitted or open-ended compositional communication outside our species.

tion once they had emerged (see Supporting Information for the full set of languages in the experiment). The difference between these two experimental settings is simply that the second introduces a new adaptive challenge for the evolving language. To be faithfully transmitted from generation to generation, a language in this experiment must be both learnable and unambiguous. The learnability constraint is imposed by the participants in the experiment, and the ambiguity constraint is imposed by our additional filter.

The result is the evolution of exactly the type of structure that optimises both these two competing constraints: compositionality. This reveals a key feature of cultural transmission: it gives rise to adaptive systems that respond to the pressures imposed by the transmission bottleneck that exists between the producer and learner of behaviour. Crucially, this is adaptation **by the language** that maximises its own transmissibility, and it can happen without intentional design on the part of the individuals involved. Participants could not be aware in the second experiment that ambiguous signals were being filtered, and yet a completely different sort of structure emerged. This demonstrates that adaptation can be independent of the intentions of individuals.

Finally, the difference between the two experiments also shows that the languages that emerge are not simply a reflection of the native language of the participants. If participants were somehow stamping their own linguistic knowledge onto the data that they were seeing, there would be no reason why in the first experiment we would find rampant structured underspecification, and in the second a system of morphological concatenation.

Conclusions

We have shown that it is possible to study cumulative cultural adaptation in the laboratory. Using a diffusion-chain paradigm with an artificial language learning task, we provide empirical support for computational and mathematical models of iterated learning that show language to be an adaptive system in its own right. We demonstrate, for the first time, the cumulative evolution of novel adaptive structure without intentional design on the part of the participants in the experiment.

We can understand the linguistic structure emerging in these experiments as an adaptive response by language to the problem of being transmitted from generation to generation. In particular, language faces the problem of being reproducible from a sub-sample. In the first experiment, the language solves this problem by introducing systematic underspecification in the meaning-signal mapping. In the second experiment, the language faces the additional challenge of being transmitted despite filtering for ambiguity. Compositional structure is a potential solution to this particular transmission problem, and this structure emerges. It is important to reiterate that participants in the experiment did not intentionally design this solution – indeed, they were not even aware of the problem. Participants believed they were reproducing as best they could the language they were exposed to. Just as biological evolution can deliver the appearance of design without the existence of a designer, so too can cultural evolution.

Methods

80 participants were recruited to participate in an “alien language” learning study. Each had to learn a language made up of written labels for visual stimuli. Participants were university students with no background in linguistics. Female:male ratio was 46:34, mean age: 22.5, minimum age: 18, maximum age: 40. The experiment was conducted in accordance with the ethics procedures of the department of Linguistics and English Language at the University of Edinburgh. Participants car-

ried out the experiment at a computer terminal and received written and verbal instructions (see Supporting Information). During training, participants were presented with string-picture pairs on the computer monitor. During testing, participants were presented with pictures on the monitor and are prompted to enter strings using the keyboard, with any sequence of alphanumeric characters being permissible.

Visual stimuli: There were 27 possible stimuli to be labeled. Each was a coloured object with an arrow indicating motion. Each object feature (shape, colour, motion) varied over three possible values: square, circle or triangle; black, blue or red; and horizontal, bounce or spiral.

Labels: The set of labels in the initial language were generated and assigned randomly and were constructed by concatenating between 2 and 4 syllables (without spaces between) taken from a set of 9 simple consonant-vowel pairs. Given that participants were free to enter any sequence of characters they chose during testing, subsequent labels were not constrained in this way.

Training and testing regime: Each language (a set of 27 string-picture pairs, one string for each of 27 possible pictures) was divided randomly into two sets: the SEEN set (14 string-picture pairs) and the UNSEEN set (13 string-picture pairs). Each participant acquired their language in a single session, comprising of three ‘rounds’ of training with an optional two-minute break in between. A single round of training consisted of two randomised exposures to the SEEN set, followed by a test. In the first two rounds this test phase contained only half of the SEEN and half the UNSEEN items; the final test at the end of the third round (which was the only source for the next generations language) consisted of all 27 pictures.

During each training pass through the SEEN set, participants were presented with each pair in a random order, with the string being displayed for 1 second, followed by both string and picture being displayed for a further 5 seconds. During testing, participants were presented with a picture and prompted to type in what they think the alien would produce for that picture.

In the second experiment, the SEEN set was filtered before presentation to participants. Specifically, if any string labels more than one picture, all but one of those string-picture pairs (chosen at random) is moved into the UNSEEN set. As a result, the training data seen by participants in the second experiment consisted of a purely one-to-one mapping from strings to pictures even if the language of the previous generation included one-to-many mappings.

Diffusion chain design: The first participant in the experiment was trained on a language with randomly-constructed labels. Subsequent participants were trained on the output of the final testing of the previous participant: the previous participant’s final testing output was randomly re-divided into a new SEEN and UNSEEN set.

Measure of transmission error: The mean distance between all the signals in a participant’s output and the corresponding signals in the previous generation’s output gives a measure of inter-generation transmission error, and is given by

$$E(i) = \frac{1}{|M|} \sum_{m \in M} LD(s_i^m, s_{i-1}^m) \quad [1]$$

where s_i^m is the string associated with meaning m by participant at generation i , $LD(s_i^m, s_j^m)$ is the normalised Levenstein distance between strings s_i^m and s_j^m , and the sum is over a set of meanings M of magnitude $|M|$.

Measure of structure: For a particular language, a measure of structure is computed as follows. The distances between all pairs of strings in the language are calculated using normalised Levenshtein distance. In addition, the distances between all pairs of meanings are also calculated using a simple hamming distance (so that meanings differing in one feature have a distance of 1, meanings differing on two features have a distance of 2 and so on). The Pearson's product-moment correlation between these two sets of distances is then calculated, giving an indication of the extent to which similar meanings are expressed using similar strings. In order to be able to compare across different languages, and to measure significance, it is necessary to compute a Monte Carlo sample of this measure under permutations of the strings over meanings. The graphs shown in the paper give the z-score for the veridical correlation based on 1000 randomisations. The dotted line on the graph therefore shows the 95% confidence interval that the observed mapping could be obtained by random assignment of signals to meanings. This measure is undefined when there is no variation in the Monte Carlo sample, for example where the language only has the same string for all meanings, or for all but one of the meanings. In these cases, all possible re-orderings are equally structured.

Acknowledgements

Cornish is funded by an ESRC 1+3 Studentship. Smith was funded by a BA Postdoctoral Fellowship.

References

1. Maynard Smith J, Szathmari E (1997) *The Major Transitions in Evolution* (Oxford Univ Press, Oxford).
2. Deacon TW (1997) *The Symbolic Species: The Co-evolution of Language and the Brain* (WW Norton, New York).
3. Christiansen MH, Chater N (2008) Language as shaped by the brain. *Behav Brain Sci* in press.
4. Kirby S (1999) *Function, Selection and Innateness: the Emergence of Language Universals* (Oxford Univ Press, Oxford).
5. Hurford JR (2000) in *The Evolutionary Emergence of Language: Social function and the origins of linguistic form*, eds Knight C, Studdert-Kennedy M, Hurford JR (Cambridge Univ Press, Cambridge), pp 324-352.
6. Kirby S, Smith K, Brighton H (2004) From UG to universals: linguistic adaptation through iterated learning. *Stud Lang* 28:587-607.
7. Kirby S, Dowman M, Griffiths T (2007) Innateness and culture in the evolution of language. *Proc Natl Acad Sci* 104:5241-5245.
8. Kirby S (2000) in *The Evolutionary Emergence of Language: Social function and the origins of linguistic form*, eds Knight C, Studdert-Kennedy M, Hurford JR (Cambridge Univ Press, Cambridge), pp 303-323.
9. Brighton H (2002) Compositional syntax from cultural transmission. *Artif Life* 8:25-54.
10. Zuidema W (2003) in *Advances in Neural Information Processing Systems 15 (Proceedings of NIPS'02)*, eds Becker S, Thrun S, Obermayer K (MIT Press, Cambridge, MA), pp 51-58.
11. Smith K, Brighton H, Kirby S (2003) Complex systems in language evolution: the cultural emergence of compositional structure. *Adv Complex Syst* 6:537-558.
12. Brighton H, Smith K, Kirby S (2005) Language as an evolutionary system. *Physics of Life Reviews* 2:177-226.
13. Griffiths TL, Kalish ML (2007) Language evolution by iterated learning with Bayesian agents. *Cognit Sci* 31:441-480.
14. Keller R (1994) *On language change: The invisible hand in language* (Routledge, London).
15. Batali J (2002) in *Linguistic Evolution through Language Acquisition: Formal and Computational Models*, eds Briscoe EJ (Cambridge University Press, Cambridge), pp 111-172.
16. Bickerton D (2007) Language evolution: a brief guide for linguists. *Lingua* 117:510-526.

17. Bartlett FC (1932) *On Remembering* (Macmillan, Oxford).
18. Curio E, Ulrich E, Vieth W (1978) Cultural transmission of enemy recognition: one function of avian mobbing. *Science* 202:899–901.
19. Laland KN, Plotkin HC (1990) Social learning and social transmission of foraging information in Norway rats (*Rattus norvegicus*). *Anim Learn Behav* 18:246–251.
20. Galef BG Jr., Allen C (1995) A new model system for studying animal traditions. *Anim Behav* 50:705–717.
21. Laland KN, Williams K (1997) Shoaling generates social learning of foraging information in guppies. *Anim Behav* 53:1161–1169.
22. Mesoudi A, Whiten A (2004) The hierarchical transformation of event knowledge in human cultural transmission. *J Cognit Cult* 4:1–24.
23. Mesoudi A, Whiten A, Dunbar RIM (2006) A bias for social information in human cultural transmission. *Br J Psychol* 97:405–423.
24. Horner V, Whiten A, Flynn E, de Waal FBM (2006) Faithful replication of foraging techniques along cultural transmission chains by chimpanzees and children. *Proc Natl Acad Sci* 103:13878–13883.
25. Gomez RL, Gerken L (2000) Infant artificial language learning and language acquisition. *Trends Cognit Sci* 4:178–186.
26. Hudson-Kam C, Newport E (2005) Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Lang Learn Dev* 1:151–195.
27. Galantucci B (2005) An experimental study of the emergence of human communication systems. *Cognit Sci* 29:737–767.
28. Garrod S, Doherty G (1994) Conversation, co-ordination and convention: an empirical investigation of how groups establish linguistic conventions. *Cognition* 53:181–215.
29. Garrod S, Fay N, Lee J, Oberlander J, MacLeod T (2007) Foundations of representation: Where might graphical symbol systems come from? *Cognit Sci* 31:961–987.
30. Selten R, Warglien M (2007) The emergence of simple languages in an experimental coordination game. *Proc Natl Acad Sci* 104:7361–7366.
31. Landau B, Smith LB, Jones SS (1988) The importance of shape in early lexical learning. *Cognit Dev* 3:299–321.
32. Slobin DI (1977) in *Language Learning and Thought*, ed Macnamara J (Academic Press, London), pp 185–221.
33. Smith K (2004) The evolution of vocabulary. *J Theoret Biol* 228:127–142.
34. Von Frisch K (1974) Decoding the language of the Bee. *Science* 185:663–668.
35. Zuberbühler K (2002) A syntactic rule in forest monkey communication. *Anim Behav* 63:293–299.

Figure captions

Figure 1: An example string-picture pair.

Figure 2: Transmission error and a measure of structure by generation in four chains. 2a shows the increase in learnability (decrease in error) of languages over time. 2b shows structure in the languages increasing. The dotted line on 2b gives the 95% confidence interval such that any result above this line demonstrates that there is a non-random alignment of signals and meanings. In other words, structure in the set of signals reflects structure in the set of meanings. In two cases, this measure is not defined and is therefore not plotted (see Methods). The example language discussed in the paper is circled.

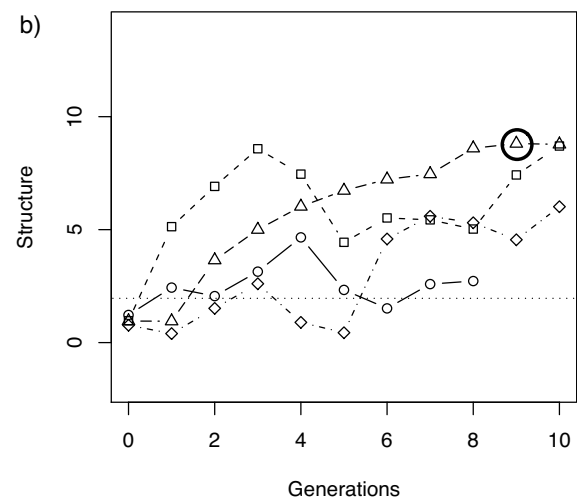
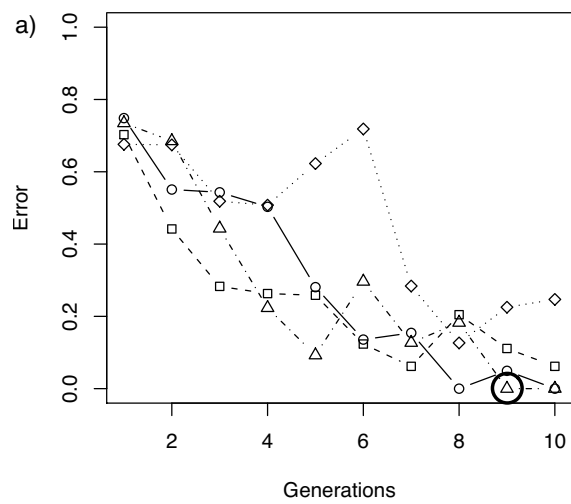
Figure 3: An example evolved language in the first experiment. This language exhibits systematic underspecification, enabling learners to reproduce the whole language from a fragment.

Figure 4: Transmission error and structure by generation, for the experiment where ambiguous data was removed from the training set at each generation. 4a gives error for the whole language, 4b gives structure. These results show that, despite the blocking of underspecification, structure still evolves which enables the languages to become increasingly learnable. The example language discussed in the paper is circled.

Figure 5: An example evolved language in the second experiment. The language is structured: the string associated with a picture consists of substrings expressing colour, shape and motion respectively. The hyphens represent one way of analysing the substructure of these strings and are added purely for clarity – participants in the experiment always produced strings of characters without spaces or any other means of indicating substructure.

kihemiwi







tuge
tuge
tuge

tuge
tuge
tuge

tuge
tuge
tuge



tupim
miniku
tupin

tupim
miniku
tupin

tupim
miniku
tupin

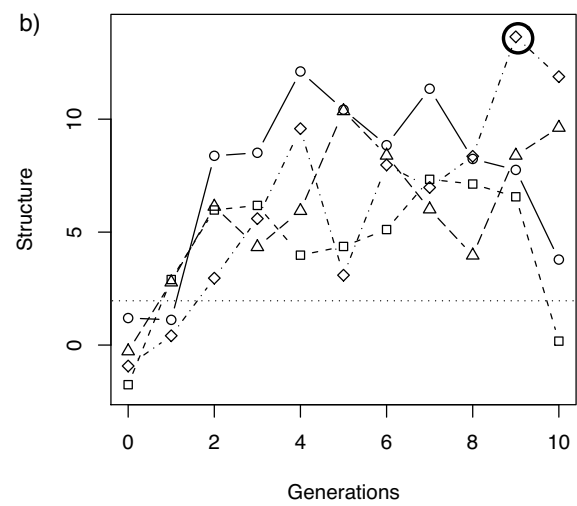
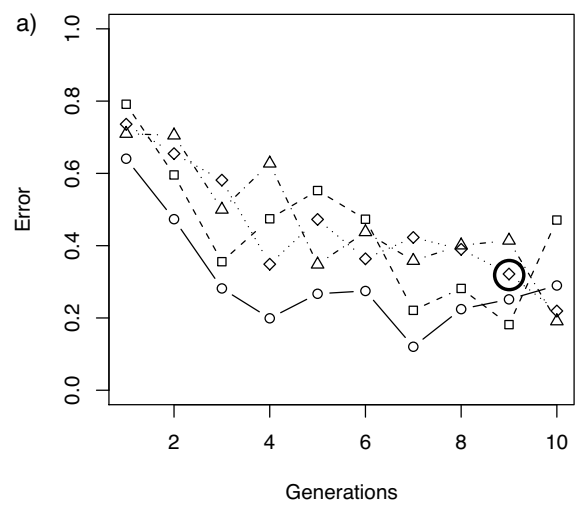


poi
poi
poi

poi
poi
poi

poi
poi
poi





→	n-ere-ki	l-ere-ki	renana	□
	n-ehe-ki	l-aho-ki	r-ene-ki	○
	n-eke-ki	l-ake-ki	r-ahe-ki	△
↗	n-ere-plo	l-ane-plo	r-e-plo	□
	n-eho-plo	l-aho-plo	r-eho-plo	○
	n-eki-plo	l-aki-plo	r-aho-plo	△
↻	n-e-pilu	l-ane-pilu	r-e-pilu	□
	n-eho-pilu	l-aho-pilu	r-eho-pilu	○
	n-eki-pilu	l-aki-pilu	r-aho-pilu	△

Table 1: Number of distinct words by generation in the first experiment (symbols correspond to those on Fig. 2).

Generation	0	1	2	3	4	5	6	7	8	9	10
○ Chain 1	27	17	9	6	5	4	4	2	2	2	2
▣ Chain 2	27	17	15	8	7	6	6	6	5	5	4
△ Chain 3	27	24	8	6	6	5	6	5	5	5	5
◊ Chain 4	27	23	9	10	9	11	7	5	5	4	4

Table 2: Number of distinct words by generation in the second experiment (symbols correspond to those on Fig. 4).

Generation	0	1	2	3	4	5	6	7	8	9	10
○ Chain 1	27	23	22	17	21	21	17	21	25	13	16
□ Chain 2	27	26	13	10	10	16	16	12	12	13	12
△ Chain 3	27	11	16	14	12	17	14	16	20	19	12
◇ Chain 4	27	19	19	17	19	17	22	23	21	27	23